<u>AMENDMENTS TO THE CLAIMS</u>

1-21 (Canceled)

22   (Currently Amended)   A harmonic structure acoustic signal detection method <u>for</u>~~of~~ detecting a segment that includes speech, as a speech segment, from an input acoustic signal <u>which is divided into a plurality of frames with a predetermined period</u>, said <u>harmonic structure acoustic signal detection</u> method comprising:

an acoustic feature extraction step of extracting an acoustic feature in each <u>frame</u> of <u>the plurality of</u> frames into which the input acoustic signal is divided ~~at every predetermined time period~~; and

a segment determination step of evaluating <u>a</u> continuity of the <u>extracted</u> acoustic features and of determining a speech segment according to the evaluated continuity,

~~wherein in said acoustic feature extraction step, frequency transform is performed on each of the frames into which the input acoustic signal is divided at every predetermined time period, and the acoustic feature that is a value of a harmonic structure represented by a number is extracted, and~~

<u>wherein said acoustic feature extraction step includes:</u>

<u>a frequency transformation step of frequency-transforming each frame of the plurality of frames to obtain components;</u>

<u>a correlation value calculation step of dividing the components obtained through said frequency transformation step into frequency bands of a predetermined bandwidth and calculating correlation a value between components in predetermined frequency bands in different frames;</u>

<u>a weight calculation step of calculating a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and</u>

3

a harmonic structure acoustic feature extraction step of extracting the acoustic feature that is a value of a harmonic structure represented by a number, using a product of the correlation value calculated in said correlation value calculating step and the weight calculated in said weight calculation step, and

wherein, in said segment determination step, the speech segment is determined based on at least one of ~~the following:~~ a correlation value between acoustic features in the same frame~~;~~ and a correlation value between acoustic features in different frames.

23-28 (Canceled)

29   (Currently Amended)   The harmonic structure acoustic signal detection method according to claim 22,

wherein, in said segment determination step, the continuity of the acoustic features is evaluated based on a correlation value between the acoustic features of different frames. ~~, and the speech segment is determined according to the evaluated continuity.~~

30   (Currently Amended)   The harmonic structure acoustic signal detection method according to claim 22,

wherein, in said segment determination step, the continuity of the acoustic features is evaluated based on distributions of the acoustic features in different frames. ~~, and the speech segment is determined according to the evaluated continuity.~~

31   (Currently Amended)   The harmonic structure acoustic signal detection method according to claim 22, further comprising:

an evaluation step of calculating an evaluation value for evaluating the continuity of the acoustic features, ~~; and~~

wherein, in said ~~a speech~~ segment determination step, the continuity evaluated is a temporal continuity. ~~of evaluating temporal continuity of the evaluation values and of determining a speech segment according to the evaluated temporal continuity.~~

4

32 **(Currently Amended)** The harmonic structure acoustic signal detection method according to claim 31,

wherein said segment determination step further includes:

a step of estimating a speech signal-to-noise ratio of the input acoustic signal <u>to be high if</u>~~based on comparisons~~, for a predetermined number of frames, ~~between (i)~~ acoustic features extracted in said acoustic feature extraction step or the evaluation values calculated in said evaluation step <u>are greater in magnitude than</u>~~and (ii)~~ a first predetermined threshold<u>,</u> ~~; and~~

~~a step of determining the speech segment~~

<u>wherein the speech segment is determined</u> based on the evaluation value calculated in said evaluation step, in the case where the estimated speech signal-to-noise ratio is ~~equal to or higher than a second predetermined threshold~~<u>estimated to be high</u>, and

<u>wherein the speech segment is determined based on an evaluated temporal continuity of the evaluation values</u>~~in said speech segment determination step, the temporal continuity of the evaluation values is evaluated and the speech segment is determined according to the evaluated temporal continuity~~, in the case where the speech signal-to-noise ratio is <u>not estimated to be high.</u> ~~lower than the second predetermined threshold.~~

33 **(Previously Presented)** The harmonic structure acoustic signal detection method according to claim 22,

wherein said segment determination step includes:

an evaluation step of calculating an evaluation value for evaluating the continuity of the acoustic features; and

a non-speech harmonic structure segment determination step of evaluating temporal continuity of the evaluation values and determining, according to the evaluated temporal continuity, a non-speech harmonic structure segment that has a harmonic structure but is not a speech segment.

34 **(Canceled)**

35   **(Canceled)**

36   **(Currently Amended)**   The harmonic structure acoustic signal detection method according to claim 22,

wherein, in said segment determination step, the continuity is evaluated based on correlation values between two or more types of frames of different time periods.

37   **(Currently Amended)**   The harmonic structure acoustic signal detection method according to claim 36,

wherein, in said segment determination step, one of the correlation values between the two or more types of frames of different time periods is selected based on a speech signal-to-noise ratio of the input acoustic signal, and the continuity is evaluated based on the selected correlation value.

38   **(Currently Amended)**   The harmonic structure acoustic signal detection method according to claim 22,

wherein, in said segment determination step, the continuity is evaluated based on a corrected correlation value calculated using a difference between (i) a correlation value between the acoustic features of frames and (ii) an average value of the correlation values of a predetermined number of frames.

39   **(Currently Amended)**   A harmonic structure acoustic signal detection device <u>for detecting</u><s>which detects</s> a segment that includes speech, as a speech segment, from an input acoustic signal <u>which is divided into a plurality of frames with a predetermined period</u>, said <u>harmonic structure acoustic signal detection</u> device comprising:

an acoustic feature extraction unit operable to extract an acoustic feature in each <u>frame of the plurality of</u> frames into which the input acoustic signal is divided <s>at every predetermined time period</s>; and

6

a segment determination unit operable to evaluate a continuity of the extracted acoustic features, and to determine a speech segment according to the evaluated continuity,

~~wherein said acoustic feature extraction unit is operable to perform frequency transform on each of the frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic structure represented by a number, and~~

wherein said acoustic feature extraction unit includes:

a frequency transformation unit operable to frequency-transform each frame of the plurality of frames to obtain components;

a correlation value calculation unit operable to divide the components obtained through said frequency transformation unit into frequency bands of a predetermined bandwidth and to calculate a correlation value between components in predetermined frequency bands in different frames;

a weight calculation unit operable to calculate a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and

a harmonic structure acoustic feature extraction unit operable to extract the acoustic feature that is a value of a harmonic structure represented by a number, using a product of the correlation value calculated in said correlation value calculating unit and the weights calculated in said weight calculation unit, and

wherein said segment determination unit is operable to determine the speech segment based on at least one of ~~the following:~~ a correlation value between acoustic features in the same frame~~;~~ and a correlation value between acoustic features in different frames.

40 **(Currently Amended)** A speech recognition device for recognizing~~which recognizes~~ speech included in an input acoustic signal which is divided into a plurality of frames

with a predetermined period, said speech recognition device comprising:

an acoustic feature extraction unit operable to ~~extract an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period~~frequency-transform each frame of the plurality of frames into which the input acoustic signal is divided and to extract an acoustic feature that is a value of a harmonic structure represented by a number;

a segment determination unit operable to evaluate a continuity of the extracted acoustic features, and to determine a speech segment according to the evaluated continuity; and

a recognition unit operable to recognize speech in the speech segment determined by said segment determination unit,

~~wherein said acoustic feature extraction unit is operable to perform frequency transform on each of the frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic structure represented by a number, and~~

wherein said acoustic feature extraction unit includes:

a frequency transformation unit operable to frequency-transform each frame of the plurality of frames to obtain components;

a correlation value calculation unit operable to divide the components obtained through said frequency transformation unit into frequency bands of a predetermined bandwidth and to calculate a correlation value between components in predetermined frequency bands in different frames;

a weight calculation unit operable to calculate a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and

a harmonic structure acoustic feature extraction unit operable to extract the acoustic feature that is a value of a harmonic structure represented by a number, using a product

8

of the correlation value calculated in said correlation value calculating unit and the weight calculated in said weight calculation step, and

wherein said segment determination unit is operable to determine the speech segment based on at least one of the following: a correlation value between acoustic features in the same frame, and a correlation value between acoustic features in different frames.

41 **(Currently Amended)** A speech recording device for recording which records speech included in an input acoustic signal which is divided into a plurality of frames with a predetermined period, said speech recording device comprising:

an acoustic feature extraction unit operable to extract an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period frequency-transform each frame of the plurality of frames into which the input acoustic signal is divided and to extract an acoustic feature that is a value of a harmonic structure represented by a number;

a segment determination unit operable to evaluate a continuity of the extracted acoustic features, and to determine a speech segment according to the evaluated continuity; and

a recording unit operable to record the input acoustic signal in the speech segment determined by said segment determination unit,

wherein said acoustic feature extraction unit is operable to perform frequency transform on each of the frames into which the input acoustic signal is divided at every predetermined time period, and to extract the acoustic feature that is a value of a harmonic structure represented by a number, and

wherein said acoustic feature extraction unit includes:

a frequency transformation unit operable to frequency-transform each frame of the plurality of frames to obtain components;

a correlation value calculation unit operable to divide the components obtained through said frequency transformation unit into frequency bands of a predetermined bandwidth and to calculate a correlation value between components in predetermined frequency bands in different frames;

9

a weight calculation unit operable to calculate a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and

a harmonic structure acoustic feature extraction unit operable to extract the acoustic feature that is a value of a harmonic structure represented by a number, using a product of the correlation value calculated in said correlation value calculating unit and the weight calculated in said weight calculation unit, and

wherein said segment determination unit is operable to determine the speech segment based on at least one of ~~the following:~~ a correlation value between acoustic features in the same frame~~;~~ and a correlation value between acoustic features in different frames.

42 **(Currently Amended)** A computer-readable recording medium storing a computer program ~~which causes~~for causing a computer to execute:

an acoustic feature extraction step of ~~extracting an acoustic feature in each of frames into which the input acoustic signal is divided at every predetermined time period~~frequency-transforming each frame of the plurality of frames into which the input acoustic signal is divided and extracting an acoustic feature that is a value of a harmonic structure represented by a number; and

a segment determination step of evaluating a continuity of the extracted acoustic features and of determining a speech segment according to the evaluated continuity,

~~wherein in said acoustic feature extraction step, frequency transform is performed on each of the frames into which the input acoustic signal is divided at every predetermined time period, and the acoustic feature that is a value of a harmonic structure represented by a number is extracted, and~~

wherein said acoustic feature extraction step includes:

a frequency transformation step of frequency-transforming each frame of the

10

plurality of frames to obtain components;

a correlation value calculation step of dividing the components obtained through said frequency transformation step into frequency bands of a predetermined bandwidth and calculating a correlation value between components in predetermined frequency bands in different frames;

a weight calculation step of calculating a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and

a harmonic structure acoustic feature extraction step of extracting the acoustic feature that is a value of a harmonic structure represented by a number, using a product of the correlation value calculated in said correlation value calculating step and the weight calculated in said weight calculation step, and

wherein in said segment determination step, the speech segment is determined based on at least one of ~~the following:~~ a correlation value between acoustic features in the same frame~~;~~ and a correlation value between acoustic features in different frames.

43   (New)   The harmonic structure acoustic signal detection method according to claim 22,

wherein said weight calculation step includes:

a band number calculation step of calculating a band number which indicates a difference between an identifier of a frequency band having a maximum value and an identifier of a frequency band having a minimum value in the correlation value in a same frame or between adjacent frames;

a corrected band number calculation step of calculating, based on a distribution of band numbers, corrected band numbers of the band numbers; and

a weighted band number calculating step of calculating a weighted band number

11

as the weight, the weighted band number being a maximum value of the corrected band numbers.

44 **(New)** A harmonic structure acoustic signal detection method for detecting a segment that includes speech, as a speech segment, from an input acoustic signal which is divided into a plurality of frames with a predetermined period, said harmonic structure acoustic signal detection method comprising:

an acoustic feature extraction step of extracting an acoustic feature in each frame of the plurality of frames into which the input acoustic signal is divided; and

a segment determination step of evaluating a continuity of the extracted acoustic features and of determining a speech segment according to the evaluated continuity,

wherein said acoustic feature extraction step includes:

a frequency transformation step of frequency-transforming each frame of the plurality of frames to obtain components;

a correlation value calculation step of dividing the components obtained through said frequency transformation step into frequency bands of a predetermined bandwidth and calculating a correlation value between components in predetermined frequency bands in the same frame;

a weight calculation step of calculating a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of the correlation values and a minimum value of the correlation values is larger than a threshold value, being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold;

a correlation value calculation step of dividing the components obtained through said frequency transformation step into frequency bands of a predetermined bandwidth, and of calculating a correlation value between the components in predetermined frequency bands in the same frame; and

an extraction step of extracting, as the acoustic feature, an identifier of a frequency band in which the component has a maximum value or a minimum value of the correlation values in the same frame,

12

wherein said segment determination step includes:

an evaluation step of calculating an evaluation value for evaluating the continuity of the acoustic features; and

a non-speech harmonic structure segment determination step of evaluating temporal continuity of the evaluation values and determining, according to the evaluated temporal continuity, a non-speech harmonic structure segment that has a harmonic structure but is not a speech segment, and

wherein, in said segment determination step, the speech segment is determined based on at least one of a correlation value between acoustic features in the same frame and a correlation value between acoustic features in different frames.

45 **(New)** A harmonic structure acoustic signal detection method for detecting a segment that includes speech, as a speech segment, from an input acoustic signal which is divided into a plurality of frames with a predetermined period, said harmonic structure acoustic signal detection method comprising:

an acoustic feature extraction step of extracting an acoustic feature in each frame of the plurality of frames into which the input acoustic signal is divided; and

a segment determination step of evaluating a continuity of the extracted acoustic features and of determining a speech segment according to the evaluated continuity,

wherein said acoustic feature extraction step includes:

a frequency transformation step of frequency-transforming each frame of the plurality of frames to obtain components;

a correlation value calculation step of dividing the components obtained through said frequency transformation step into frequency bands of a predetermined bandwidth and calculating a correlation value between components in predetermined frequency bands in frames which are a predetermined number of frames away from each other;

a weight calculation step of calculating a weight, in a same frame or between adjacent frames, the calculated weight, when a difference between a maximum value of the correlation values and a minimum value of the correlation values is larger than a threshold value,

13

being smaller than the calculated weight when the difference between the maximum value of the correlation values and the minimum value of the correlation values is smaller than the threshold; and

an acoustic feature extraction step of extracting the acoustic feature that is a value of a harmonic structure represented by a number, by calculating a distribution of the correlation values in every predetermined number of frames, and

wherein, in said segment determination step, the speech segment is determined based on at least one of a correlation value between acoustic features in the same frame and a correlation value between acoustic features in different frames.